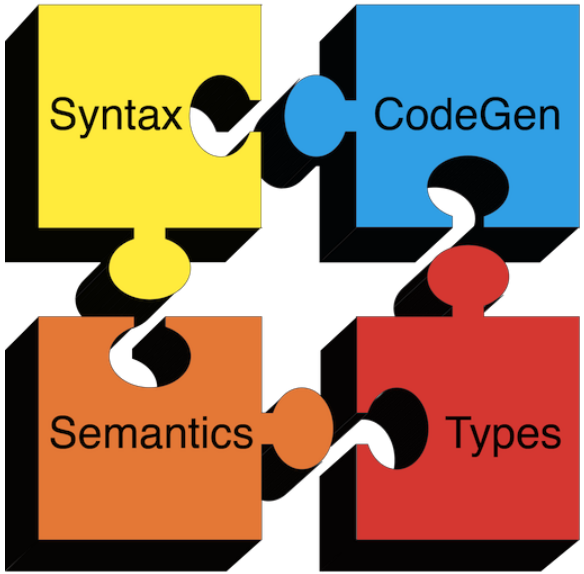

Compilers

Lexical Specifications

Lexical Specifications

Keyword: “if” or “else” or “then” or …

Lexical Specifications

Integer: *a non-empty string of digits*

Lexical Specifications

Identifier: *strings of letters or digits, starting with a letter*

Whitespace: *a non-empty sequence of blanks, newlines, and tabs*

Lexical Specifications

anyone@cs.stanford.edu

Lexical Specifications

$$
\begin{aligned}
\text{digit} &= \text{'0'} + \text{'1'} + \text{'2'} + \text{'3'} + \text{'4'} + \text{'5'} + \text{'6'} + \text{'7'} + \text{'8'} + \text{'9'} \\
\text{digits} &= \text{digit}^{+} \\
\text{opt_fraction} &= (\text{'.'}\ \text{digits}) + \varepsilon \\
\text{opt_exponent} &= (\text{'E'}\ (\text{'+'} + \text{'-'} + \varepsilon)\ \text{digits}) + \varepsilon \\
\text{num} &= \text{digits opt_fraction opt_exponent}
\end{aligned}
$$

Lexical Specifications

Choose the regular languages that are correct specifications of the English-language description given below:

Twelve-hour times of the form "04:13PM". Minutes should always be a two digit number, but hours may be a single digit.

- ☐ (0 + 1)?[0-9]:[0-5][0-9](AM + PM)
- ☐ ((0 + ε)[0-9] + 1[0-2]):[0-5][0-9](AM + PM)
- ☐ (0*[0-9] + 1[0-2]):[0-5][0-9](AM + PM)
- ☐ (0?[0-9] + 1(0 + 1 + 2):[0-5][0-9](A + P)M

Lexical Specifications

- Regular expressions describe many useful languages
- Regular languages are a language specification
  - We still need an implementation
- Next time: Given a string *s* and a rexp *R*, is

$$s \in L(R)?$$